Clustering Household Electricity Use Profiles

Conference Paper · December 2013

DOI: 10.1145/2542652.2542656

1 author:

John Richard Williams
University of Otago
30 PUBLICATIONS **768** CITATIONS

Clustering Household Electricity Use Profiles

John Williams
Department of Marketing
University of Otago
Dunedin, New Zealand
john.williams@otago.ac.nz

ABSTRACT

An attempt was made to cluster the load profiles of a sample (n≈ 380) of New Zealand households. An extensive range of approaches was evaluated, including the approach of clustering on "features" of the data rather than the raw data. A semi-automatic search of the problem space (cluster base, distance measure, cluster/partitioning method and k) resulted in a $k = 3$-cluster solution with acceptable quality indices and face validity. Although a particular combination of base, distance metric and clustering method was found to work well in this case, it is the practice of searching the problem space, rather than a particular solution, that is discussed and advocated.

Categories and Subject Descriptors

G.3 [**Probability and Statistics**]: Time Series Analysis

General Terms

Algorithms, Performance, Experimentation

Keywords

Clustering, Time series, Distance metric, Electricity

1. INTRODUCTION

This paper describes an attempt to cluster electricity load profiles (*i.e.* the amount of electricity used over a certain period, *e.g.* day, hour, or smaller time period). To do this, it is necessary to find a way to compare two vectors (ordered sets) of the same length such that a difference between them can be defined in a meaningful way. The difference in question relates to the *shape* of the vectors when plotted, *e.g.* whether the maxima and minima appear at the same or similar positions.

This is an important problem for both engineering and financial reasons. The national electricity distribution grid needs to be able to withstand peak loads, so if it were found (for example) that a cluster with excessive peak loads corresponds with a household demographic that was expected to increase in the coming years, this would provide an early warning sign. Also, for business reasons electricity retailers have a pressing need to shift peak loads, *i.e.* to encourage consumers to use less electricity at peak times, because during those times they are selling electricity to consumers for a lower price than what they pay to electricity producers (*i.e.* the "wholesale" price of electricity varies during the day, but the retail price does not).

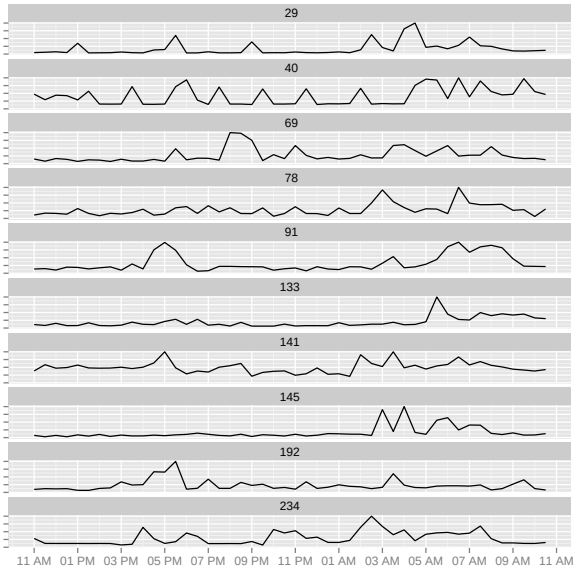

Figure 1: Half-hourly electricity usage of 10 households over one day

To make things more clear, example data is shown in Figure 1. The purpose of this figure is to illustrate the difficulty of judging the similarity between even a small number of load profiles. While the data that were clustered consists of ≈ 380 households, ideally for commercial application any clustering approach should be able to deal with thousands or tens of thousands of vectors.

2. PRIOR RESEARCH

Traditional clustering algorithms treat the each observation of a set of variables as *unordered* sets, *i.e.* any permutation of the order in which the variables appear in the data

MLSDA '13, December 3, 2013, Dunedin, New Zealand

file will produce identical results. For longitudinal data this obviously is a severe limitation. Normal clustering methods, whether hierarchical methods like nearest neighbour and Ward's method [23], or partitioning methods like k-means [14] or k-medians, typically produce miserable results when used on longitudinal data.

Although this issue has been recognised for a few years, clustering longitudinal data is not a solved problem, in general. Rather it is an active area in both the statistics and machine learning literatures. One promising approach, known as "Characteristic-Based clustering" (CBC) [21], is not to cluster the data themselves, but rather "features" of the data, *i.e.* quantities calculated from the data, *e.g.* mean, IQR, % used at peak times and so on. The attraction of this approach is that (ideally) it captures the essential properties of each observation, but with fewer data per case. This is important for two reasons. Firstly, many cluster analysis methods do not perform well when variables are correlated. With time-series data, this is obviously a problem. Secondly, it helps to reduce the "curse of dimensionality" that is common to many multivariate methods, especially cluster analysis and factor analysis. Finally, clustering on features means that the method can be applied to sets of time series of differing lengths and/or with missing values.

When developing CBC, the researchers used the following 13 features in their work [21]:

- Measures calculated on the raw data
    1. serial correlation
    2. non-linearity
    3. skewness
    4. kurtosis
    5. self-similarity
    6. chaotic
    7. periodicity
- Measures calculated on de-trended and seasonally adjusted data
    1. trend
    2. seasonality
    3. serial correlation
    4. non-linearity
    5. skewness
    6. kurtosis

At first glance it might seem that some of these features might not apply to electricity profiles measured over the period of one day, in particular trend and seasonality. While it is common in feature-based classification to calculate measures that assume symmetric or even Gaussian distribution, it is not clear that it is particularly valid to do so. In particular, the distribution of values of kWh over almost any time period (day, week, month, year) is not even approximately Gaussian distributed, and nor is there any reason to suspect or assume that it should be. But *whatever works*, regardless of whether it is theoretically relevant, is the prime criterion for selecting a method of clustering the data at hand.

The features were all normalised before cluster analysis, as were the input (raw) data. They used SOM (self-organising maps, [13] to cluster their data. However they only assessed the quality of their method by classification accuracy (*i.e.* where the true assignment of cases to clusters is known).

The same researchers, later working in the context of univariate time-series forecasting [22], used the same features but estimated them in a slightly different way. They have had numerous requests for the **R** code that extracts these features, and so they made the code available from the web site of one of the authors (Rob Hyndman, a well-known expert in forecasting).[1]

The idea of clustering on features has been used by other researchers to cluster electricity data. For example, researchers attempting to cluster household electricity use in Finland [16] followed the CBC approach[21] but used different features. They analysed the electricity usage of 1035 households over 84 days, sampled at hourly intervals. They extracted seven features from each customer's data:

1. mean
2. standard deviation
3. skewness
4. kurtosis
5. chaos
6. energy
7. periodicity

More theoretically interesting and relevant features relate to the predictability of the data, *i.e.* periodicity and "chaos" (the term used by these authors ([21, 16]) for the (maximal) Lyapunov exponent (MLE). While the standard descriptive statistics should be familiar, perhaps the Lyapunov exponent, "energy" (as used in this context) and the way to measure periodicity may not be so familiar. Hence the definitions are presented below.

The Lyapunov exponent is defined [18] as:

$$\lambda \equiv \lim_{t\to\infty} \frac{1}{t} \ln \frac{\delta Z(t)}{\delta Z_0} \tag{1}$$

where δZ_t is the separation of two trajectories (to use the terminology in which this measure was developed) at time t, and δZ_0 is the initial separation between them.

It is well-known that the Lyapunov exponent cannot be calculated algebraically, rather numerical methods must be used. A number of different methods have been proposed, and most require the user to select particular parameters that are specific to the data and theoretical problem at hand. Fortunately, the **R** code made available by Rob Hyndman provides an MLE measure that is not computationally expensive.

The other measure that might be unfamiliar is how "Energy" is defined in this particular context:

$$\text{Energy} \equiv \frac{\sum_{i=1}^{w} |x_i|^2}{w} \tag{2}$$

where x_i are the FFT components of a time window of length w. In periodogram analysis w is usually taken to be $T/2+1$,

[1] http://robjhyndman.com/hyndsight/tscharacteristics/

where T is the length of the vector in question. Periodicity was assessed using the Discrete Power Spectrum [15]

Selection of k was performed by calculating the Davies-Bouldin Index (DBI) [5] for each value of k and then selecting the solution with the highest DBI value. Goodness of the resulting solution was assessed using the Index of Association [24]:

$$\mathrm{IA} \equiv 1 \frac{\sum_{i=1}^{n}(P_i - O_i)^2}{\sum_{i=1}^{n}(|P_i'| + |O_i'|)^2} \tag{3}$$

where $P_i' \equiv P_i - \bar{O}$ and $O_i' \equiv O_i - \bar{O}$ (and P_i are the "predicted" values — *i.e.* the cluster centroids, O_i the observed values and $\bar{O}$ is the mean of the observed values.

The Finnish researchers showed that by using k-means to cluster on the seven features listed above, they were able to improve the mean IA from 0.30 (where the "predicted" values were provided by the electricity company) to 0.65 [16].

3. SAMPLE

The data to be analysed are a sample of approximately 380 households located in a suburb of Auckland, New Zealand (Pakuranga) and have been previously reported on in the context of consumer response to time-varying electricity tariffs [19]. Data are available at half-hourly intervals over a period of two years (with some missing values). Following best-practice recommendations [4], data for one weekday (in this case, Wednesdays) was averaged (using the median) over a period of three months (the first quarter of 2008). This has the effect of reducing the heterogeneity of the data, and providing a more accurate reflection of the typical daily load pattern for each household by "averaging out" one-off or rare conditions (*e.g.* visiting relatives, a snow-storm, or a power outage).

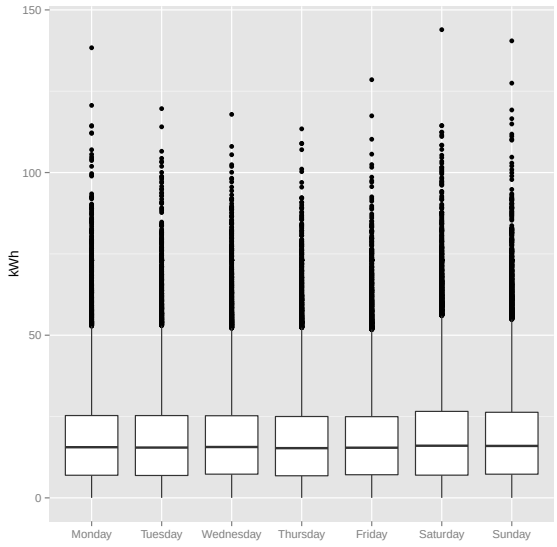

Figure 2: Daily electricity consumption, by Weekday

Figure 2 shows the median electricity usage for each of the 389 households in the sample over all the weekdays from 1 August 2007 to 31 July 2009. Note that while the median daily usage is around 15 kWh during the week (and 16 kWh on weekends), the maxima are *much* higher: about 113–143 kWh, almost an order of magnitude greater than the median. In any case, the point of this figure is to show that typical usage over one day does not vary much by day of the week, so the choice of which weekday to analyse should not matter too much.

4. METHODS AND RESULTS

This section describes the process of finding the best overall method of forming the clusters and choosing the most useful value of k. It is essentially a 3D problem space, as follows:

1. The base variables on which to cluster the profiles:
   (a) raw data or features
   (b) if features, which ones?
2. The method of forming the clusters, including decisions on:
   (a) which algorithm/method
   (b) distance metric for methods which can use more than one
3. How many clusters to retain, especially:
   (a) Which measures to use in order to evaluate the quality of the result
   (b) How to decide on k in the case of conflicting indications from multiple measures

Each of these issues are discussed below.

4.1 Methods

4.1.1 *Analysis strategy*

The overall analysis strategy was as follows:

1. Investigate which methods perform best in general, by testing them over various values of k on different bases (data, and combinations of features). This was achieved by use of the **R** packages clValid and clusterSim and a package developed for this project (in order to overcome the limitations of the other packages), called clusterTS.
2. Once the best method(s) were identified, investigate various values of k more carefully.

4.1.2 *Base variables*

The following features were used to implement the CBC approach:

- The 13 variables used by originators of CBC [21, 22]
- The four additional variables used by the Finnish researchers [16] (mean, standard deviation, energy and periodicity)
- Additional variables chosen for this research because they are important to this particular application:
  – median electricity used (over the course of one day)

- total electricity used (over the course of one day)
- proportion of electricity used at peak times
- discrete wavelet transforms (DWT) of varying levels

For CBC, this gives $13 + 4 + 3 = 20$ base variables (not including the DWT variables), a substantial reduction from the 48 variables that comprise the raw data.

Clustering on the base data was also performed, as was using hierarchical clustering on the distance matrix of the base data, with various distance metrics. The metrics used are those provided in the **R** package TSclust:

- ACF: autocorrelation
- COR: correlation
- CRT: temporal correlation
- DWT: Discrete Wavelet Transform, with the degree chosen automatically
- DTW: Dynamic Time Warping
- ECL: Euclidean distance
- EC2: Squared Euclidean distance
- IPR: Integrated Periodogram [12]
- PCF: Partial ACF
- PER: Periodogram

4.1.3 Cluster validation methods

A great number of ways of assessing the quality of clustering results have been proposed and discussed in the literature (see [9], for example). The most widely accepted strategy is not to rely on a single index, but rather several. But this leads to another problem: how to decide which index to trust when several offer conflicting indications? It is for this reason that many applications only use a single index.

However the **R** package clValid provides several classes of quality measures, and an automatic way of ranking the best cluster solution based on the information provided by several quality indices concurrently. However such "automatic" ranking should only be used as a guide as it is certainly not valid in all possible cases.

The most important indices of clustering quality (or goodness of fit) for the problem at hand are the "internal" measures, *i.e.* those that depend only on the data and the partition(s) of the data. A brief description of the indices used is given below.

1. The silhouette value [17], which indicates the certainty with which a case "belongs" in a cluster. Values of 1 indicate that the observation definitely belongs in the cluster to which it is assigned, and -1 indicates that it definitely does not. Hence values close to 1 are desirable and values close to or less than 0 are undesirable.
2. The Dunn index: range is $[0, \infty)$ and higher values are better [6]
3. MIA: Median Index of Agreement: ranges from 0 to 1 and higher values are better [24].
4. Caliński-Harabasz: higher values are better [3].
5. Davies-Bouldin: *lower* values are better [5]. This should be carefully noted when inspecting the figures and tables below.

The values of these five measures for clustering on the scaled data for a range of k between 2 and 20 are shown in Figure 5 in the appendix.

4.1.4 Choosing the clustering methods

Standard practice for clustering.

The fundamental problem facing the data analyst who is trying to find clusters in data is how to choose the most appropriate method of clustering. A second problem is how to choose the number of clusters. A very common practice is to perform hierarchical clustering and interpret the results of this (usually by visual examination of a dendrogram) in order to choose k, the number of clusters. There are several hierarchical methods of forming clusters (single-linkage, medioids, *etc.*) but Ward's method is probably the most often used. Although hierarchical cluster analysis *may* be used to assign cases to clusters, often it is only used to choose the value of k (or a set of values). Then, some form of partitioning scheme (non-hierarchical cluster analysis) is used to assign cases to cluseters. By far the most popular algorithm is k-means, or some variation (trimmed k-means, k-medians, k-means++, *etc.*). So, for the vast majority of cluster analysis applications, the analyst produces a dendrogram (usually using Euclidean or squared Euclidean distance in combination with Ward's method), chooses k (or possibly a set of plausible values of k) and then assigns cases to clusters using k-means. Some very careful analysts realise that k-means is sensitive to starting conditions, so they repeat their analyses several times with different seeds, to satisfy themselves regarding the "stability" of the clusters (*i.e.* the property that the same cases are allocated to the same cluster in subsequent runs). This is regarded as the "state of the art" (or at least "standard practice") in some University courses on statistics.

Automatic searches.

But in the current era, this approach is not acceptable, at least not for situations in which the risk of producing a "wrong" results is high and the consequences severe, for example in medical research. Instead, modern applications harness the advances in computing power to explore a decision space of clustering methods (where the space is 2D or 3D: method and k, or measure, method and k) and employ a range of quantitative measures with which to judge the quality of the result. In some applications it is possible to more-or-less automatically choose measure, method and k with a single function call.

As previously metioned, clValid [2] provides such facilities, and hence it was used for the problem at hand. An especially appealing feature of this package is that it provides a number of ways of choosing the "optimal" combination of clustering method and choice of k. In particular:

1. Methods: hierarchical, kmeans, diana, fanny, som, model, sota, pam, clara, and agnes. (For hierarchical meth-

ods, agglomeration options available are ward, single, complete, and average.)

2. Distance: euclidean, correlation, and manhattan.
3. Scaling: none
4. Quality indices: Silhouette and Dunn (and others not used for this problem)

The clusterSim package [20] was also used, because it offers some features that clValid does not have (and vice versa). Importantly, it provides a much wider array of automatic comparisons of methods, but can only evaluate the performance of each using a single index at a time. This package performs searches over the following dimensions:

1. Method: single linkage, complete linkage, average linkage, McQuitty, pam, Ward's method, centroid and median.
2. Distance: Bray-Curtis, Canberra, Chebyschev, Euclidean, GDM1 [10], Manhattan, Squared Euclidean
3. Scaling: n1: (x-mean) / sd), n2: (x-Me) / MAD, n3: (x-mean) / range, n4: (x-min) / range, n5: (x-mean) / max[abs(x-mean)], n6: (x / sd), n7: (x / range), n8: (x / max), n9: (x / mean), n10: (x / sum), n11: x / sqrt(SSQ).
4. Quality indices: Caliński & Harabasz; Baker & Hubert; Hubert & Levine; Silhouette; Krzanowski & Lai.

For $k = 2$ to 20, this gives a total of $6,992$ combinations that were searched. In this application, the ranking of models was evaluated using the [3] and Silhouette indices (meaning that 13,984 separate clustering solutions were produced and compared). After an extensive comparison of these options (details available on request) it was found that pam (**p**artitioning **a**round **m**edoids [11]) and Ward's method were clearly superior, $x/\max$ and $x/\bar{x}$ were the best scaling methods and squared Euclidean the best distance methods. Because $x/\max$ is more intuitively interpretable this was chosen as the most appropriate scaling method. It was difficult to decide between Ward's method and pam, but where there was a difference between them pam was almost always superior. Unfortunately it is not possible to automatically compare the hierarchical clustering methods with partitioning methods using clusterSim, so clValid was used for this purpose.

4.1.5 *Choosing the value(s) of k*

Building on the exploratory results of running clValid and clusterSim, clusterTS was used on the base variables listed in Section 4.1.2, with three hierarchical methods (agnes, pam on the distance matrix and diana), and three partitioning methods (k-means, pam, and SOM). These six methods were evaluated for a k ranging from 2 to 20. Agnes and diana are the hierarchical methods provided in the cluster package. Agnes is agglommerative (**ag**golmerative **nes**ting) and diana is divisive (**di**visive **ana**lysis).

Additionally, kml [8], which is k-means for longitudinal data, and apcluster [1], which implements Affinity Propagation Clustering [7] were used (via my own package clusterTS) to investigate particularly interesting values of k. These packages were not used in the main runs because kml has a bug due to a conflict with another package used in the main run, and AP clustering is computationally expensive, making it unsuitable for automatic searching.

4.2 Results

k	C-H	Dunn	D-B	Sil	MIA	Base	Methods	Metric
Sorted by Caliński-Harabasz								
3	106.66	0.18	1.63	0.15	0.79	data	kmeans	ECL
3	106.60	0.18	1.74	0.15	0.79	data	som	ECL
3	99.35	0.15	1.36	0.15	0.77	data	pam	ECL
4	87.01	0.16	1.42	0.15	0.82	data	kmeans	ECL
4	86.86	0.17	1.88	0.15	0.82	data	som	ECL
Sorted by Dunn								
11	48.46	0.22	1.08	0.13	0.88	data	som	ECL
12	45.94	0.22	1.15	0.13	0.88	data	som	ECL
9	54.19	0.21	1.13	0.13	0.87	data	som	ECL
8	57.85	0.20	1.15	0.14	0.86	data	som	ECL
6	68.30	0.19	1.94	0.13	0.85	data	kmeans	ECL
Sorted by Davies-Bouldin								
5	26.33	0.09	0.40	0.13	0.80	data	diana	COR
6	23.32	0.09	0.41	0.12	0.81	data	diana	COR
7	20.07	0.09	0.41	0.12	0.81	data	diana	COR
11	11.70	0.09	0.48	0.25	0.80	data	diana	ACF
12	11.21	0.09	0.50	0.22	0.80	data	diana	ACF
Sorted by median IA								
20	33.96	0.18	1.15	0.11	0.90	data	som	ECL
20	33.66	0.16	1.82	0.12	0.90	data	kmeans	ECL
19	35.42	0.17	1.04	0.13	0.90	data	som	ECL
19	35.30	0.16	1.85	0.12	0.89	data	kmeans	ECL
18	36.35	0.19	0.78	0.12	0.89	data	som	ECL
Sorted by Silhouette								
3	19.59	0.09	3.89	0.71	0.73	data	diana	IPR
4	13.76	0.09	1.36	0.70	0.73	data	diana	IPR
3	**23.82**	**0.07**	**4.79**	**0.69**	**0.74**	**data**	**agnes**	**IPR**
5	10.64	0.08	1.37	0.65	0.73	data	diana	IPR
6	14.17	0.07	1.37	0.65	0.75	data	diana	IPR

Table 1: The five best models by each quality index

The results, shown in Table 1 and Figure 5 in the Appendix, are mixed. Things we can learn from this information include:

- The Caliński-Harabasz and silhouette indices indicate that fewer clusters are better, while the Dunn, Davies-Bouldin and MIA indicate the opposite. There is an inherent trade-off between fitting the data closely (*i.e.* producing a high index of agreement) and the certainty that a particular case belongs to the cluster to which it is assigned, and not a neighbouring cluster. The more clusters there are, the more likely it is that the centroid of one will be close to the centroid of another.
- Most of these measures vary fairly linearly with k, except the Dunn and Davies-Bouldin indices.
- The Dunn and Davies-Bouldin indices vary with k in a very different manner than do the other indices. The usual advice when using these indices is to look for an "elbow", but there appear to be no distinct elbows in these plots.

We can see from the figure that despite its theoretical weakness for longitudinal data, Euclidean distance on the data, not the features, provides better values on most or all indices.

While the silhouette and IA measures have absolute maxima (unity in both cases), the other measures have no cut-off values for an "acceptable" solution. The general method of using all these indices (even those with theoretical maxima) is to use the "elbow" or "scree" criterion, familiar to users of exploratory factor analysis. Taken together, these results indicate that either the minimum or maximum value of k are "optimal". This is not an encouraging sign!

Regarding the value of k, this is a decision that, for this particular application, should be based on not only statistical criteria, but also managerial relevance. For example, 20 market segments may be far too many for any meaningful marketing strategy, and two or three far too few.

Table 1, which shows the five best models as defined by each quality index, summarises the results in terms of the problem space defined by the decision criteria listed above. There are a number of pertinent facts that we can glean from this table:

- In all cases the best fit is achieved when the base is the data, not the features.
- In the majority of cases Euclidean distance gives the best results. But this is not as clear-cut as the indication that the (scaled) raw data is the superior base.
- The Integrated Periodogram (IPR) distance measure, coupled with hierarchical clustering, produces clusters that are best on the silhouette measure and also have relatively large median IAs. Closer inspection of these results produced by diana revealed that this was because they allocated the majority of cases to one or two large clusters, with the remaining clusters consisting of only four cases in one cluster or two clusters consisting of two cases each.

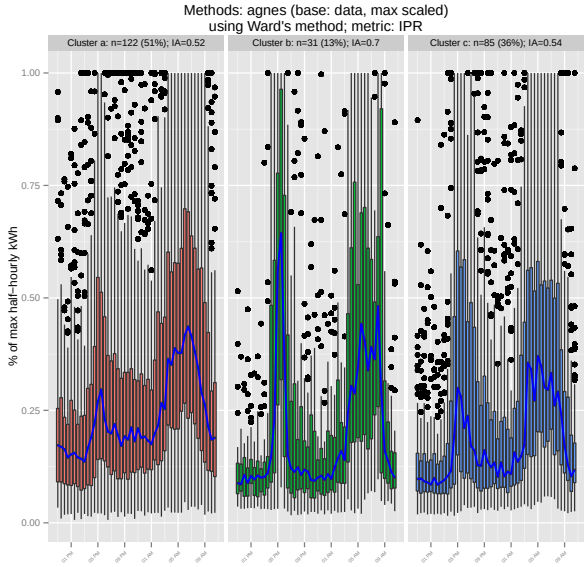

Figure 3: Graphical summary of dispersion of observations around centroids

Inspections of the best fitting models by the IA criterion revealed a similar problem to the best fitting models by silhouette: the mean or median fit was inflated by very good fit in clusters with only a handful of observations. With $k = 20$ and $n \approx 380$, and the well-known propensity of k-means to produce clusters of roughly equal size, this meant that a few clusters had fewer than 10 cases allocated to them.

However using agnes to cluster on the IPR distance measure produces acceptable results, in the sense that usual criterion of minimising within-cluster variance (measured by median IA) while maximising between-cluster variances (measured by the silhouette value, albeit indirectly) is met. In addition is produces plausible cluster sizes: not all similar as k-means tends to produce, nor with clusters with a very small proportion of cases. Hence, this is the cluster solution that will be retained and explored.

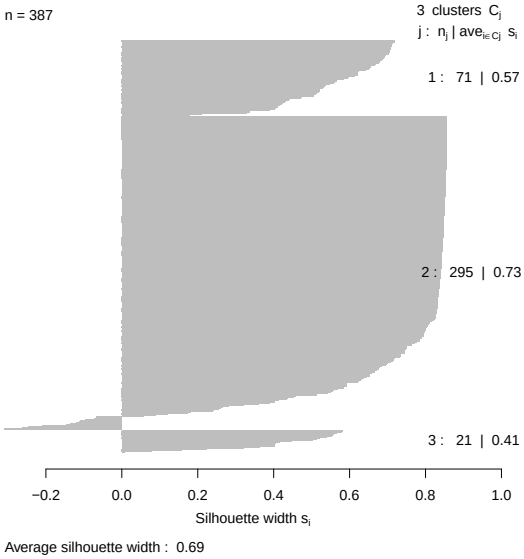

Figure 4: Silhouette plot of the $k = 3$ cluster solution

		Indices of Agreement					
Cluster	Size	Mean	SD	Median	IQR	Min	Max
1	71	0.76	0.12	0.79	0.12	0.40	0.91
2	295	0.70	0.13	0.72	0.15	0.28	0.93
3	21	0.79	0.11	0.79	0.16	0.52	0.92

Table 2: Indices of agreement for the $k = 3$ cluster solution

Figure 3 shows the dispersion of cases around the centroids of each cluster and Figure 4 show the silhouette plot. We can see from the information in Figure 4 that the clusters are well separated and with centroids that make sense intuitively: the largest cluster (b, 76%) is comprised of household that use more electricity in the evenings than in the mornings, the second largest (a, 18%) uses approximately equal amounts of electricity in the morning and evening, while the smallest cluster (c, 5%) uses significantly more electricity in the morning than the evening.

These usage patterns correspond to plausible household types, particularly with respect to how much cooking, water heating, space heating and use of consumer electronics (televisions, home theaters, gaming consoles *etc.*) occurs in the evenings. Cluster c may correspond to single-person, rel-

atively high-income households who eat out frequently, for example.

5. CONCLUSION

An extensive examination of bases on which to cluster the load profiles, ways to measure the distance between load profiles and algorithms to partition the data resulted in the choice to use agglomerative hierarchical clustering on the integrated periodogram distance measure between the scaled raw profiles (rather than features calculated from the profiles).

This runs somewhat counter to some published research that advocate features-based clustering, but in *this particular application* the result is perhaps not so surprising. Two of the advantages of features-based clustering, dimension reduction and dealing with missing data, do not apply here. Because the data were the median usage averaged over several days there were no missing values. And because there were only 48 dimensions in the raw data and over 20 in some of the feature-based models, the dimension reduction, though large proportionally can be considered not so large in absolute terms.

But the major message of this paper is this: although in this case hierarchical clustering with IPR distance performed well, it is by no means claimed that this combination will work well in general, or that CBC does not live up to the claims of those who advocate it.

What *is* indicated by these results, however, is the need to explore a range of bases, distance measures, paritioning algorithms and cluster quality indices in a systematic manner, rather than searching for a "magic bullet" algorithm that works well for all data of a similar nature to that presented here.

6. REFERENCES

[1] Ulrich Bodenhofer, Andreas Kothmeier, and Sepp Hochreiter. Apcluster: an R package for affinity propagation clustering. *Bioinformatics*, 27:2463–2464, 2011.

[2] Guy Brock, Vasyl Pihur, Susmita Datta, and Somnath Datta. clvalid: An R package for cluster validation. *Journal of Statistical Software*, 25(4):1–22, 3 2008.

[3] T. Caliński and J. Harabasz. A dendrite method for cluster analysis. *Communications in Statistics*, 3(1):1–27, 1974.

[4] Gianfranco Chicco. Overview and performance assessment of the clustering methods for electrical load pattern grouping. *Energy*, 42:68–80, 2012.

[5] D. Davies and D. Bouldin. A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2:224–227, 1979.

[6] J.C. Dunn. Well separated clusters and fuzzy partitions. *Journal of Cybernetics*, 4:95–104, 1974.

[7] Brendan J. Frey and Delbert Dueck. Clustering by passing messages between data points. *Science*, 315:972–977, 2007.

[8] Christophe Genolini. *kml: K-means for Longitudinal data*, 2012. R package version 2.1.2.

[9] Maria Halkidi, Yannis Batistakis, and Michalis Vazirgiannis. On clustering validation techniques. *Journal of Intelligent Information Systems*, 17:107–145, 2001.

[10] K. Jajuga, M. Walesiak, and A. Bak. On the general distance measure. In M. Schwaiger and O. Opitz, editors, *Exploratory data analysis in empirical research*, pages 104–109. Springer-Verlag, Berlin, Heidelberg, 2003.

[11] Leonard Kaufman and Peter J. Rousseeuw. *Finding Groups in Data*. Wiley Series in Probability and Statistics. Wiley, New York, 1990.

[12] Claudia Klüppelberg and Thomas Mikosch. The integrated periodogram for stable processes. *The Annals of Statistics*, 24(5):1855–1879, October 1996. Mathematical Reviews number (MathSciNet): MR1421152; Zentralblatt MATH identifier: 0898.62116.

[13] Tuevo Kohonen. *Self-Organizing Maps*. Springer Series in Information Sciences. Springer, Berlin, 2nd edition, 1997.

[14] J. B. MacQueen. Some methods for classification and analysis of multivariate observations. In L. M. LeCam and J. Neyman, editors, *Prodeedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 281–297, 1967.

[15] T. Masters. *Neural, Novel & Hybrid Algorithms for Time Series Prediction*. John Wiley & Sons Inc., New York, 1995.

[16] Teemu Räsänen and Mikko Kolehmainen. Feature-based clustering for electricity use time series data. In Mikko Kolehmainen, Pekka Toivanen, and Bartlomiej Beliczynski, editors, *Adaptive and Natural Computing Algorithms*, volume 5495 of *Lecture Notes in Computer Science*, pages 401–412. Springer Berlin Heidelberg, 2009.

[17] Peter J. Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65, 1987.

[18] J.C. Sprott. *Chaos and Time-Series Analysis*. Oxford University Press, Oxford, 2003.

[19] Paul Thorsnes, John Williams, and Rob Lawson. Consumer responses to time varying prices for electricity. *Energy Policy*, 49:552–561, 2012.

[20] Marek Walesiak and Andrzej Dudek. *clusterSim: Searching for optimal clustering procedure for a data set*, 2013. R package version 0.42-1.

[21] Xiaozhe Wang, Kate Smith, and Rob Hyndman. Characteristic-based clustering for time series data. *Data Mining and Knowledge Discovery*, 13:335–364, 2006.

[22] Xiaozhe Wang, Kate Smith-Miles, and Rob Hyndman. Rule induction for forecasting method selection: meta-learning the characteristics of univariate time series. *Neurocomputing*, 72:2581–2594, 2009.

[23] Joe H. Ward. Hierarchical grouping to optimize an objective function. *Journal of American Statistical Association*, 48:236–244, 1963.

[24] C. Willmot. Some comments on the evaluation of model performance. *Bulletin of American Meteorological Society*, 63:1309–1313, 1982.

7. APPENDIX

(see over page)

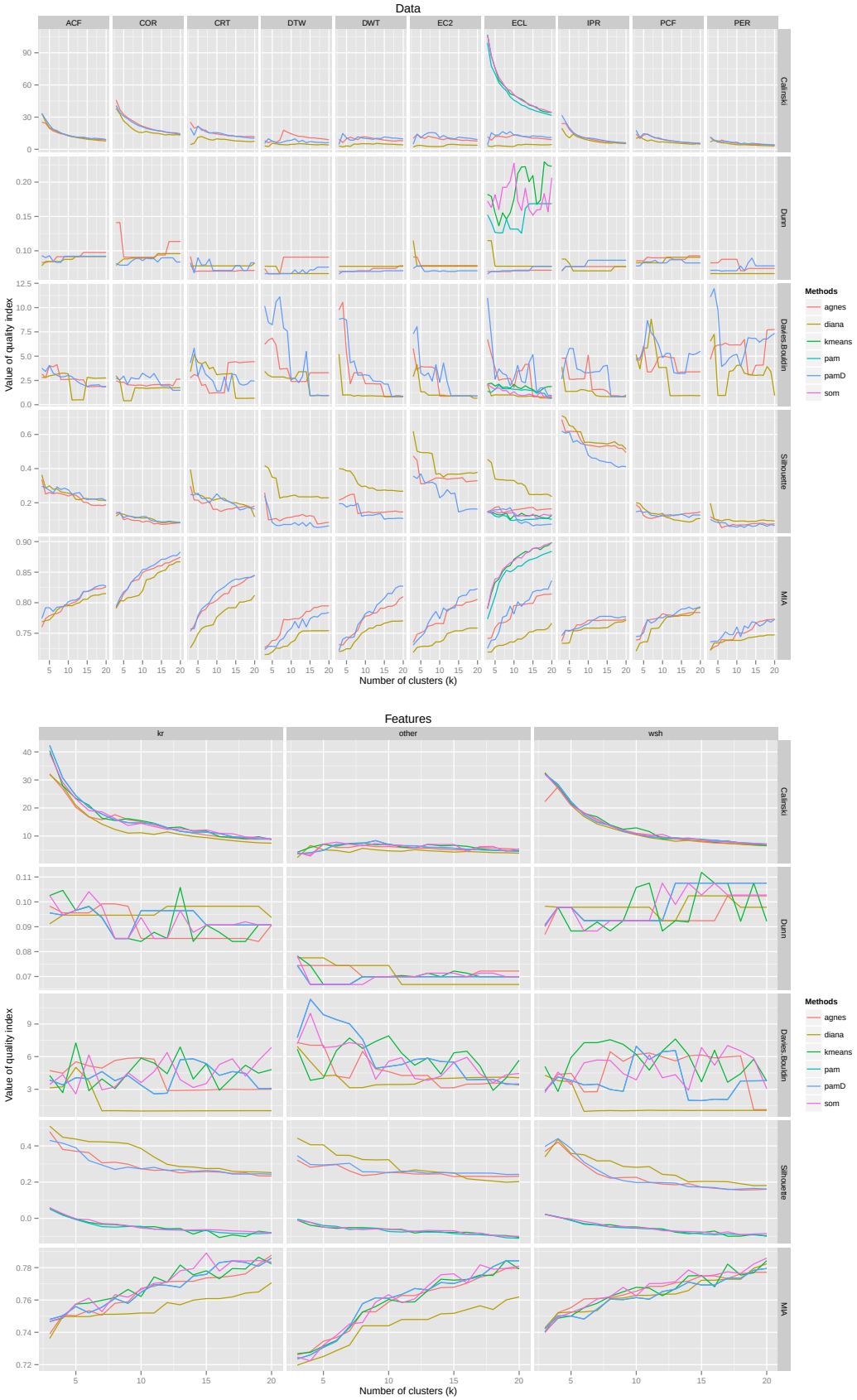

Figure 5: Quality indices for 2 <= k <= 20, agnes, diana, k-means, pam, pam on distance matrices, and SOM. For Caliński-Harabasz, Dunn, Silhouette and MIA, higher values are better; for Davies-Bouldin lower values are better.